UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Natural Language Processing

2425-2-FDS01Q011

Obiettivi

Lo scopo del corso è fornire un'introduzione ai concetti fondamentali relativi all'elaborazione del linguaggio naturale (NLP) nonché una
panoramica dei principali strumenti utilizzati nel settore. Inoltre, verranno presentate alcune applicazioni di NLP, ad es. information retrieval, traduzione automatica e rilevamento di discorsi di odio.

Contenuti sintetici

Il contenuto del corso include i principi fondamentali dell'elaborazione del linguaggio naturale (NLP) e offre una panoramica degli strumenti chiave utilizzati in questo campo. Il corso coprirà una vasta gamma di argomenti, che vanno dalle tecniche statistiche ai recenti progressi negli approcci neurali. Inoltre, il corso comprende dimostrazioni pratiche di diverse applicazioni di NLP, tra cui information retrieval, traduzione automatica e rilevamento dell'incitamento all'odio.

Programma esteso

Intoduzione al corso

- Rationalist and Empiricist Approaches to Language
- The Ambiguity of Language: Why NLP Is Difficult
  Linguistic Essentials
- Lexical resources
- Zipf’s laws
- Collocations

- Concordances
- Syntax
  Frequentist Representation of Text (TF, TF-IDF, etc..) e Word Embeddings
- Word2Vec
- FastText
- Glove
  Techiche di visualizzazione di embeddings:
    - Principal Components Analysis
- T-distributed stochastic neighbor embedding
- Uniform Manifold Approximation and Projection
  Sequence-to-Sequence (RNN, LSTM)
  Transformers and Large Language Models
- Attention Mechanisms: Self and Multi Head Attention
  Contextualized Language Models:
- ELMO
- BERT
- GPT
- LLAMA
  Prompting and Instruct Tuning
  Transformers and Large Language Models
- Interpretability and Explainability of Language Models

Prerequisiti

Basic knowledge of statistics and programming languages.

Modalità didattica

Il corso sarà tenuto in lingua inglese e si articolerà sia in lezioni frontali che introdurranno gli argomenti principali, sia in sessioni tutoriali in cui verranno spiegati gli strumenti open source.
Possono far parte del corso seminari tenuti da esperti a livello nazionale e internazionale.

12 lezioni da 2 ore svolte in modalità erogativa in presenza;
12 lezioni da 2 ore svolte in modalità interattiva (lezioni asincrone).

Materiale didattico

Daniel Jurafsky and James Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2008.

Emily M. Bender, "Linguistic Fundamentals for Natural Language Processing", Synthesis lectures on human language technologies, Morgan&Claypool Publishers, 2013.

Yoav Goldberg, "Neural Network Methods for Natural Language Processing", Synthesis lectures on human language technologies, Morgan&Claypool Publishers, 2017.

Mohammad Taher Pilehvar and Jose Camacho-collados, "Embeddings in Natural Language Processing",

Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2021.

Periodo di erogazione dell'insegnamento

First Semester

Modalità di verifica del profitto e valutazione

Progetto
• Il progetto consiste nello sviluppo di uno strumento per l'elaborazione del linguaggio naturale basato su metodi e modelli presentati durante il corso.
• E' necessario identificare un dominio di interesse e un set di dati per il quale si intende affrontare compiti specifici che coinvolgano metodi, modelli e strumenti di NLP.
• Il progetto deve essere presentato oralmente
• Il progetto viene valutato nell'intervallo [0-24]

Esame orale
• La prova orale può avere un esito compreso tra [-8; +8]
• Si compone di 4 domande su argomenti affrontati durante il corso
-2 punti per una risposta errata o per nessuna risposta, +2 punti per una risposta corretta.

Orario di ricevimento

To be agreed with the teacher

Sustainable Development Goals